FEBS Letters 539 (2003) 2–6

FEBS 27063

Hypothesis

A novel strategy to design binding molecules harnessing the modular nature of repeat proteins

Patrik Forrer, Michael T. Stumpp, H. Kaspar Binz, Andreas Plückthun*

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 9 January 2003; accepted 7 February 2003

First published online 3 March 2003

Edited by Gianni Cesareni

Abstract **Repeat proteins, such as ankyrin or leucine-rich repeat proteins, are ubiquitous binding molecules, which occur, unlike antibodies, intra- and extracellularly. Their unique modular architecture features repeating structural units (repeats), which stack together to form elongated repeat domains displaying variable and modular target-binding surfaces. Based on this modularity, we developed a novel strategy to generate combinatorial libraries of polypeptides with highly diversified binding specificities. This strategy includes the consensus design of self-compatible repeats displaying variable surface residues and their random assembly into repeat domains. We envision that such repeat protein libraries will be highly valuable sources for novel binding molecules especially suitable for intracellular applications.**

Key words: Ankyrin repeat; Combinatorial library; Leucine-rich repeat; Protein design; Repeat protein; Scaffold

1. Introduction

Man-made polypeptide libraries are indispensable sources for specific binding molecules suitable for research and biomedical applications. Such libraries can be designed by choosing a suitable prototype protein as a scaffold, whose surface is then randomized by localized amino acid substitutions or insertions [1–3]. This approach is greatly facilitated by a known three-dimensional structure of the prototype protein, which allows one to rationally select substitution or insertion points. A protein scaffold chosen for the derivation of a library of potential binders should ideally possess all of the following properties: a large and diverse target-binding surface, high thermodynamic stability under oxidizing and reducing conditions, high stability on storage, efficient folding and expression properties, and low immunogenicity.

Currently, by far the most widely used protein scaffolds are antibody fragments, i.e. single-chain Fv and Fab fragments [4,5]. The potential target-binding surface of these molecules is largely formed by the complementarity determining region loops. Antibody fragments are able to mediate high-affinity and very specific interactions to small and large molecules [6,7]. Nevertheless, they can have limitations in expression yield and stability, especially under reducing conditions, as would be encountered in intracellular applications [8,9]. Other successfully applied scaffolds include the Z domain of staphylococcal protein A (affibodies) [10], *Escherichia coli* thioredoxin [11,12], staphylococcal nuclease [13], lipocalins [14], green fluorescent protein [15], and fibronectin type III domains [16,17]. Binding molecules based on these scaffolds, with the exception of affibodies, are conceptually similar to antibodies: they have functionally partitioned protein architectures consisting of a structural framework and variable target-binding surface loops. In contrast, affibodies achieve their binding function through variable amino acids distributed over a flat surface formed by two adjacent α-helices of the structural framework [10].

In all these scaffolds, the binding surface is limited by the size of the scaffold. Repeat proteins [18,19], on the other hand, have evolved another successful binding strategy. They feature repeating structural units (repeats), which stack together to form elongated protein domains (repeat domains) with a continuous target-binding surface, which is variable in size as the number of repeats can be varied [20–22]. Residues on the surface of secondary structure elements and in loops can, depending on the type of repeat, contribute to the interaction surface. Each repeat contributes both to the stability of the domain and to the potential target-binding surface of the domain. We present here a novel strategy to generate binding molecules, which is based on the modularity of repeat proteins.

2. Evaluation of repeat proteins

Repeat proteins constitute, next to immunoglobulins, the most abundant natural protein classes specialized in binding [18,19,23]. They are found in all phyla, they occur intra- and extracellularly and they are involved in diverse biological processes, such as cell cycle control, transcriptional regulation, innate immunity, vesicular trafficking, cell differentiation, apoptosis, cellular scaffolding or bacterial invasion.

2.1. The modular nature of repeat proteins

Repeat proteins feature consecutive copies of small (about 20–40 amino acid residues) structural units (repeats) stacking together to form repeat domains (Figs. 1 and 2D) [20–22]. Such repeats have a well-defined folding topology and may contain α-helices, β-strands, or both. Examples of repeats include leucine-rich repeats (LRRs), ankyrin repeats (ARs), armadillo/HEAT repeats and tetratricopeptide repeats (Table 1) [18,20–22,24,25]. In a repeat domain, only repeats of a

*Corresponding author. Fax: (41)-1-635 5712.
E-mail address: plueckthun@bioc.unizh.ch (A. Plückthun).

Abbreviations: AR, ankyrin repeat; LRR, leucine-rich repeat

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)00177-7

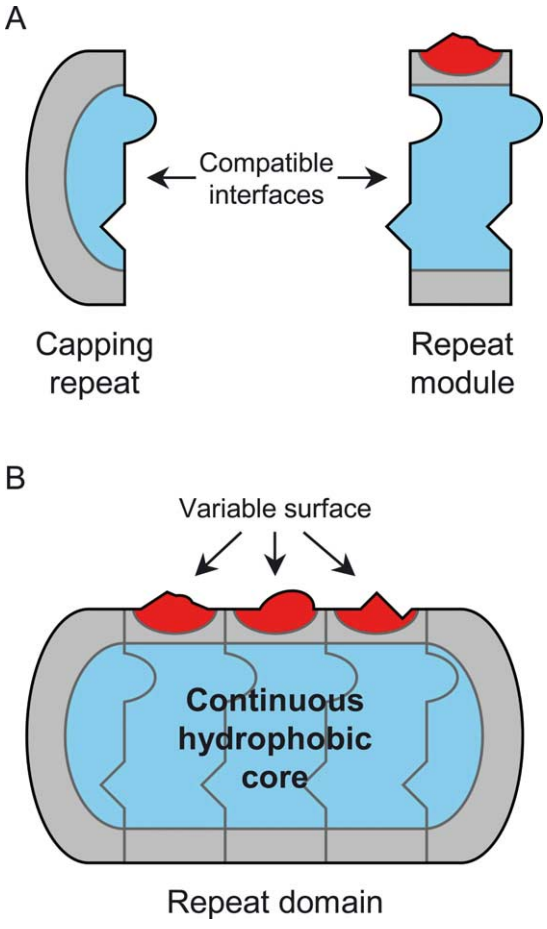

Fig. 1. Schematic representation of the designed repeat domains. A: Designed capping repeats and designed self-compatible repeat modules are the building blocks of the repeat domains. Compatible interfaces allow the stacking of repeat modules. B: The designed repeat domains consist of an N- and C-terminal capping repeat flanking a variable number of (here: three) repeat modules. The repeat modules stack together forming a continuous hydrophobic core, which is sealed on both sides by a capping repeat. The repeat domains display variable molecular surfaces, which are potential target binding sites. A,B: The hydrophobic core is shown in blue, the more hydrophilic conserved molecular surface is represented in gray, and the variable surface is depicted in red.

given type are combined, and every repeat tightly interacts with the preceding and following repeat. Capping repeats form only one such interaction, thereby shielding the hydrophobic core of the domain from the solvent and thus terminating the repeat domain (Fig. 1). The structural compatibility of the repeats within a repeat domain is achieved by conserved framework residues that mediate the essential inter-repeat interactions. The conservation of structurally important amino acids is therefore the key to the unique modular architecture of repeat domains: the conserved interfaces between repeats suggest that individual repeats can be exchanged, deleted or inserted without destroying the tertiary structure of the domain. A large number of repeats can assemble into domains, since addition of further repeats should not be spatially restricted. Variable surface residues define the functional specificity of individual repeats, and all of these residues from stacked repeats together form the target-binding surface of a repeat domain (Fig. 2E). High-affinity binding to a target can thus be achieved by the sheer size of the interaction surface. A sufficiently high rigidity of the repeat proteins prevents unnecessary loss of entropy upon binding to the desired target and may also help to decrease interactions with irrelevant targets.

In summary, natural repeat proteins are highly versatile binding molecules due to their modular architecture and variable molecular surfaces generated by the assembly of multiple compatible repeats. This principle of structural organization appears to be an economic and successful way of evolving binding molecules, which may be exploited in developing libraries of binding molecules based on designed repeat proteins.

3. Designing repeat proteins

Having analyzed the principles of structural organization of natural repeat proteins, we developed a strategy to build combinatorial libraries of repeat proteins. The idea fundamental to our strategy is to extract information from compatible

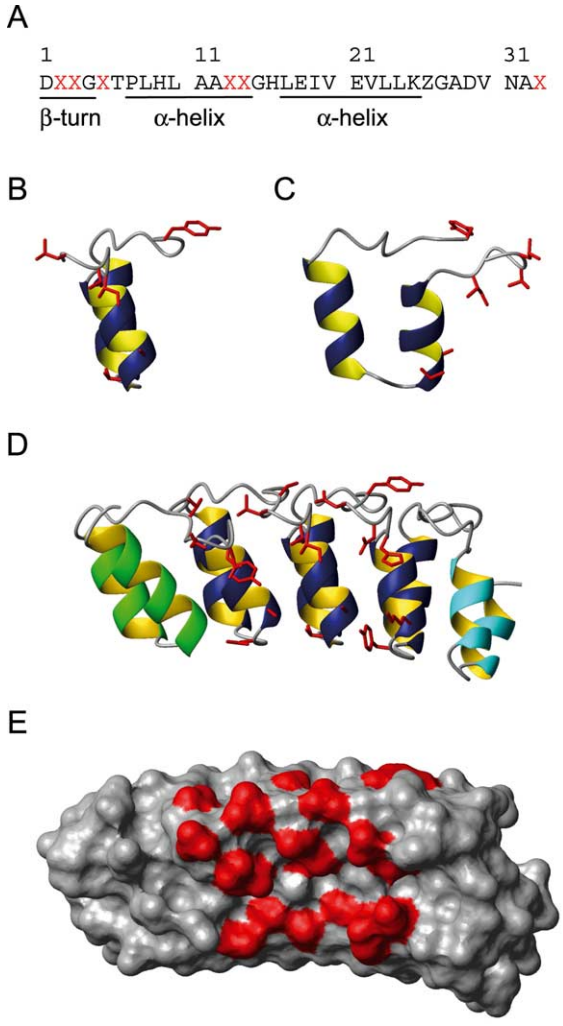

Fig. 2. Repeat sequence motif, repeat module, and X-ray structure of the designed AR protein E3_5. E3_5 is a randomly chosen member of an N3C (N- and C-terminal capping repeats flanking three designed repeat modules) AR protein library constructed using the repeat sequence motif of A (H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz and A. Plückthun, in preparation). The structure of E3_5 was solved to 2.0 Å resolution (PDB code 1MJ0) [30]. A: The designed AR sequence motif. X: Any amino acid but not G, C or P. Z: Any of the amino acids H, N, Y. B,C: Ribbon representation of the third repeat (second repeat module) of E3_5 in two perpendicular views. The orientation of the module in B corresponds to that in D. D: Ribbon representation of E3_5 showing the helices of the N-terminal capping repeat, internal repeat modules, and the C-terminal capping repeat in green, dark blue and light blue, respectively. B–D: The side chains of amino acids at variable positions are highlighted in red. E: Surface representation of E3_5 showing the conserved surface in gray and the variable surface in red. All figures were prepared with MOLMOL [33].

natural repeats to design an amino acid sequence motif encoding self-compatible repeat modules (Figs. 1A and 2A–C). Such a designed repeat sequence motif comprises fixed and variable positions. The fixed positions mainly reflect conserved framework positions of the compatible natural repeats, while the variable positions mainly reflect non-conserved surface-exposed residues that are potentially able to engage in interactions with the target. Using such a designed sequence motif, repeat modules able to stack together while displaying variable surface residues can be obtained. The conserved interfaces of the self-compatible repeat modules provide a simple means of constructing and evolving repeat proteins, as the insertion or deletion of repeat modules within a repeat domain becomes possible (Figs. 1A and 2D,E). Another important aspect of our strategy is the capping of repeat domains, which ensures that the hydrophobic module interfaces are not exposed to the solvent (Fig. 1). Such a capping may increase the folding yield and the solubility of repeat domains by making them more resistant to aggregation.

Based on these general ideas, we developed our strategy to design repeat proteins with novel binding specificities. This strategy comprises several steps: (i) the analysis of natural repeat proteins to define a set of compatible repeats, mainly by homology-based sequence searches and structural analyses; (ii) the extraction of a consensus repeat sequence motif encoding self-compatible repeat modules displaying variable surface residues by an iterative process of sequence and structural analyses; (iii) the generation of DNA fragments encoding the extracted sequence motif; (iv) the random assembly of these DNA fragments into fragments encoding distinct repeat numbers; (v) the addition of DNA sequences encoding appropriate capping repeats to the assembled DNA fragments; (vi) the translation of the assembled DNA into a library of repeat domains; and, finally, (vii) selection for particular binding specificities. The repeat sequence motif derived in step (ii) may be used to further refine the set of compatible repeat modules from step (i) by either extending or narrowing the family of homologous proteins which is considered. The ultimately derived repeat sequence motif is thus the result of an iterative process involving both the careful analysis of the input repeat database and the resulting consensus repeat sequence motif for correlating residues. We think that such a refinement further improves the quality of the designed repeat modules in terms of self-compatibility, as this process selects against the occurrence of incompatible repeat modules present in the input repeat database.

3.1. Designing self-compatible repeat modules

The crucial step of our strategy is certainly the derivation of a repeat sequence motif encoding self-compatible repeat modules. Natural repeat modules are not necessarily self-compatible, because they may have evolved together with their neighboring repeat modules leading to non-standardized module interfaces. Such evolved repeat modules are still compatible with their respective neighboring repeat modules, but have lost their self-compatibility. We thus reasoned that the careful consensus sequence analysis of compatible natural repeat modules would average out such incompatibilities and result in a consensus sequence encoding self-compatible repeat modules. Conserved positions of such a consensus sequence will include all framework residues essential for intra- and inter-repeat module interactions. Non-conserved positions may accommodate variable residues that do not restrict the self-compatibility of the repeat modules. Thus, non-conserved positions are predestined for the introduction of diversity. In natural repeats, residues at non-conserved positions often contribute to their particular biological function, i.e. target binding. Structural analyses of natural repeat domains complement the consensus design, especially for the definition of less conserved positions. Consensus design may also result in repeat proteins with improved stability and folding properties, when compared to natural ones, as anticipated from previous studies [6,26–28]. A highly stable protein is more tolerant to destabilizing mutations and such mutations may be introduced during selection procedures and may be functionally important. Fig. 2A shows such a derived repeat sequence motif obtained by implementing our strategy, which includes consensus design using the AR sequences from the SMART database [29] as initial data set (H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz, and A. Plückthun, in preparation).

3.2. Assembly of repeat modules

The modular nature of repeat proteins readily suggests building designed repeat domains by assembling repeat modules. This is best done on the DNA level by sequential assembly of DNA fragments, each encoding a diversified repeat module based on the designed repeat sequence motif. These fragments may be randomly linked, thereby resulting in sequences encoding repeat domains comprising variable numbers of repeat modules. To allow for a better-controlled assembly, we prefer to assemble the DNA fragments step-wise (M.T. Stumpp, P. Forrer, H.K. Binz and A. Plückthun, in preparation; H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz and A. Plückthun, in preparation), which results in the generation of repeat domain libraries characterized by a defined number of repeat modules (Fig. 2D,E). Insertion, shuffling or deletion of repeat modules of such repeat domains can also be performed on the DNA level, which would not be possible with other scaffolds. For example, it may be possible to elongate a selected binder by simple addition of a further library

Table 1
Repeat proteins are very abundant in nature

Repeat name	SMART[a] abbreviation	Repeats[b]	Proteins[b]	Examples	Review article
Leucine-rich repeat	LRR	14 722	2 035	Ribonuclease inhibitor [34], internalins [35]	[19]
Ankyrin repeat	ANK	7 958	1 770	IκBα [36], ankyrinR [37]	[22]
Tetratricopeptide repeat	TPR	4 883	970	$p67^{phox}$ [38], cyclophilin [39]	[40]
Armadillo/HEAT repeat	ARM	1 542	253	β-catenin [41], importin [42]	[43]

[a]SMART database: http://smart.embl-heidelberg.de/ [29].
[b]Total number in the SMART database as of 13 December 2002.

repeat module on either side. Thereby, the potential target-binding surface may be increased and thus binders with improved affinities may be obtained.

3.3. Capping of assembled repeat modules

An important feature of our strategy is that capping repeats terminate an assembly of repeat modules (Fig. 1). We think that capping repeats stabilize a repeat domain against aggregation, as they shield the hydrophobic core of repeat domains from the surrounding solvent, while an assembly devoid of capping repeats exposes an unsatisfied module interface at both sides of the stack. Capping repeats are designed to bind to those unsatisfied interfaces, thereby sealing the stack. Appropriate capping repeats can be generated either from a designed repeat module by rendering one of its interfaces more hydrophilic or by adaptation of natural capping repeats to the designed repeat module interface.

3.4. Designed LRR and AR proteins

Using our novel strategy we were able to construct highly diverse LRR and AR protein libraries (M.T. Stumpp, P. Forrer, H.K. Binz and A. Plückthun, in preparation; H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz and A. Plückthun, in preparation). Randomly chosen members from such libraries show good expression, folding and stability properties, exceeding those of corresponding natural repeat proteins, while displaying variable surface residues [30] (Fig. 2D,E). These analyses indicate that the consensus design used to create self-compatible repeat modules, the assembly of such repeat modules into repeat domains, and the capping of repeat domains were indeed successfully implemented in our strategy.

4. Conclusions and future prospects

We developed a novel strategy harnessing the modular nature of repeat proteins to build diverse polypeptide libraries, whose members display highly diversified surfaces (Fig. 2E). The idea fundamental to this strategy is to extract information from compatible natural repeats to design a repeat sequence motif encoding self-compatible repeat modules displaying variable surface residues. The key steps of our strategy are the design of an appropriate repeat sequence motif, the random assembly of repeat modules into a repeat domain, and the capping of repeat domains. So far, we have been able to successfully implement this strategy for LRR (M.T. Stumpp, P. Forrer, H.K. Binz and A. Plückthun, in preparation) and AR proteins [30] (H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz and A. Plückthun, in preparation). We envision that this strategy is also applicable to other families of repeat proteins.

Our designed repeat proteins differ in several important respects from 'classical' protein scaffolds, which rely on the presentation of a small number of more or less flexible loops. First, they are built from small units contributing both to the structural framework and to the target-binding surface. Second, the target-binding surface of repeat proteins is not dimensionally restricted as the number of repeats in a repeat domain is, per se, not limited. Last, the modular architecture of repeat domains allows the development of novel evolutionary strategies, such as module shuffling, module insertions, or module deletions. Moreover, the size of the potential target-binding surface is adaptable as desired. For example, an elongation strategy may be used for affinity maturation of selected binders.

Our repeat protein libraries were used in selections for binding against target proteins. Indeed, we isolated specific binding molecules against several globular proteins with affinities in the low nanomolar range (H.K. Binz, P. Amstutz, M.T. Stumpp, P. Forrer and A. Plückthun, unpublished) by using ribosome display [31,32]. We envision that our repeat protein libraries are highly valuable sources for novel binding molecules suitable for biotechnological and biomedical applications, and since they are designed to contain no cysteines, they may be especially suitable for intracellular or proteomics applications.

Acknowledgements: We thank Patrick Amstutz for valuable discussions and Victor Krasnykh for comments on the manuscript. M.T.S. was the recipient of a FCI and BMBF predoctoral fellowship. H.K.B. was the recipient of a predoctoral fellowship of the Roche Research Foundation. This project was supported by the NCCR Structural Biology and the Swiss Cancer Research Grant KFS 1055-09-2000.

References

[1] Ladner, R.C. and Ley, A.C. (2001) Curr. Opin. Biotechnol. 12, 406–410.
[2] Skerra, A. (2000) J. Mol. Recognit. 13, 167–187.
[3] Nygren, P.-Å. and Uhlén, M. (1997) Curr. Opin. Struct. Biol. 7, 463–469.
[4] Winter, G. (1998) FEBS Lett. 430, 92–94.
[5] Plückthun, A., Krebber, A., Krebber, C., Horn, U., Knüpfer, U., Wenderoth, R., Nieba, L., Proba, K. and Riesenberg, D. (1996) in: Antibody Engineering, A Practical Approach (McCafferty, J., Hoogenboom, H.R. and Chiswell, D.J., Eds.), pp. 203–252, Oxford University Press, New York.
[6] Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A. and Virnekäs, B. (2000) J. Mol. Biol. 296, 57–86.
[7] Söderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Åberg, A.M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., Danielsson, L., Carlsson, R. and Borrebaeck, C.A. (2000) Nat. Biotechnol. 18, 852–856.
[8] Wörn, A., Auf der Maur, A., Escher, D., Honegger, A., Barberis, A. and Plückthun, A. (2000) J. Biol. Chem. 275, 2795–2803.
[9] Biocca, S., Ruberti, F., Tafani, M., Pierandrei-Amaldi, P. and Cattaneo, A. (1995) Biotechnology 13, 1110–1115.
[10] Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M. and Nygren, P.-Å. (1997) Nat. Biotechnol. 15, 772–777.
[11] Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J. and Brent, R. (1996) Nature 380, 548–550.
[12] Cohen, B.A., Colas, P. and Brent, R. (1998) Proc. Natl. Acad. Sci. USA 95, 14272–14277.
[13] Norman, T.C., Smith, D.L., Sorger, P.K., Drees, B.L., O'Rourke, S.M., Hughes, T.R., Roberts, C.J., Friend, S.H., Fields, S. and Murray, A.W. (1999) Science 285, 591–595.
[14] Skerra, A. (2001) Rev. Mol. Biotechnol. 74, 257–275.
[15] Abedi, M.R., Caponigro, G. and Kamb, A. (1998) Nucleic Acids Res. 26, 623–630.
[16] Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L., Weng, S., Wagner, R.W. and Lipovsek, D. (2002) Chem. Biol. 9, 933–942.
[17] Koide, A., Bailey, C.W., Huang, X. and Koide, S. (1998) J. Mol. Biol. 284, 1141–1151.
[18] Andrade, M.A., Perez-Iratxeta, C. and Ponting, C.P. (2001) J. Struct. Biol. 134, 117–131.
[19] Kobe, B. and Kajava, A.V. (2001) Curr. Opin. Struct. Biol. 11, 725–732.
[20] Groves, M.R. and Barford, D. (1999) Curr. Opin. Struct. Biol. 9, 383–389.
[21] Kobe, B. and Kajava, A.V. (2000) Trends Biochem. Sci. 25, 509–515.

[22] Sedgwick, S.G. and Smerdon, S.J. (1999) Trends Biochem. Sci. 24, 311–316.
[23] Bork, P. (1993) Proteins 17, 363–374.
[24] Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 19, 415–421.
[25] Kobe, B. (1996) Nat. Struct. Biol. 3, 977–980.
[26] Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P.G.M. and Wyss, M. (2002) Protein Eng. 15, 403–411.
[27] Steipe, B., Schiller, B., Plückthun, A. and Steinbacher, S. (1994) J. Mol. Biol. 240, 188–192.
[28] Mosavi, L.K., Minor Jr., D.L. and Peng, Z. (2002) Proc. Natl. Acad. Sci. USA 99, 16029–16034.
[29] Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P. and Bork, P. (2002) Nucleic Acids Res. 30, 242–244.
[30] Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Plückthun, A. and Grütter, M.G. (2003) Proc. Natl. Acad. Sci. USA 100, 1700–1705.
[31] Hanes, J. and Plückthun, A. (1997) Proc. Natl. Acad. Sci. USA 94, 4937–4942.
[32] Amstutz, P., Forrer, P., Zahnd, C. and Plückthun, A. (2001) Curr. Opin. Biotechnol. 12, 400–405.
[33] Koradi, R., Billeter, M. and Wüthrich, K. (1996) J. Mol. Graph. 14, 51–55.
[34] Kobe, B. and Deisenhofer, J. (1995) Nature 374, 183–186.
[35] Schubert, W.-D., Urbanke, C., Ziehm, T., Beier, V., Machner, M.P., Domann, E., Wehland, J., Chakraborty, T. and Heinz, D.W. (2002) Cell 111, 825–836.
[36] Jacobs, M.D. and Harrison, S.C. (1998) Cell 95, 749–758.
[37] Michaely, P., Tomchick, D.R., Machius, M. and Anderson, R.G.W. (2002) EMBO J. 21, 6387–6396.
[38] Lapouge, K., Smith, S.M., Walker, P.A., Gamblin, S.J., Smerdon, S.J. and Rittinger, K. (2000) Mol. Cell 6, 899–907.
[39] Taylor, P., Dornan, J., Carrello, A., Minchin, R.F., Ratajczak, T. and Walkinshaw, M.D. (2001) Structure 9, 431–438.
[40] Blatch, G.L. and Lässle, M. (1999) BioEssays 21, 932–939.
[41] Huber, A.H. and Weis, W.I. (2001) Cell 105, 391–402.
[42] Vetter, I.R., Arndt, A., Kutay, U., Görlich, D. and Wittinghofer, A. (1999) Cell 97, 635–646.
[43] Andrade, M.A., Petosa, C., O'Donoghue, S.I., Müller, C.W. and Bork, P. (2001) J. Mol. Biol. 309, 1–18.